# ACADIS:

## A Research Community Data Portal

Lessons learned from a User Centered development approach

2015 GO-ESSP

Eric Nienhouse, Nathan Hook, Don Middleton

(NCAR)



CC BY

# Sample of NCAR Data Archives

*We build data archives and curate data for diverse communities.*

20K annual users, 300 data providers, 20K collections, 5PB, 2.5PB yearly downloads.




## ACADIS

Advanced Collaborative Arctic Data Information Service

NSF Arctic projects
Self publishing tools
Many disciplines
Long term preservation

## ESGF & ESG-NCAR

*Climate Data at NCAR*

Climate models (CESM)
RCMs (NARCCAP)
Large data volume
Heavily accessed

## RDA

*Research Data Archive*

Reanalysis + obs products
Subset and re-format svcs
ECMWF, ICOADS, JRA-55
Actively curated

# ACADIS: Overview

*Provide tools and services for NSF Arctic research data preservation and re-use*

*Arctic regions are changing rapidly. Research products are critical to future study.*

## Project Mission

- Support NSF Arctic community with data management leadership and services.
- Long term preservations and open source for NSF Arctic research data legacy.
- Educate and support researchers in meeting responsibilities of data preservation.
- Implement proven technologies for data preservation, access and discovery.



CC BY

# ACADIS: Overview

*Multi-disciplinary data from social science to paleoclimate and oceanography*


Many scientific domains, some with minimal data & metadata standards

# Scientific Data Preservation

*Goal: High-functioning, re-usable data by many disciplines*

Now →

| Term | Definition | Implications |
|---|---|---|
| **Curation** | **Processes that add value to foster discovery and reuse** | **Allows others to reuse data; supports cross-disciplinary research** |
| **Preservation** | **Providing enough representation information to support use and interpretation by agents other than the original data producer** | **Ability to use own data in the future (e.g., 5 yrs out)** |
| **Archiving** | **Activities that enable long-term retention of digital materials (e.g., replication, fixity, identifiers, etc.)** | **Provides identifiers for sharing, citation, etc.** |
| **Storage** | **Basic physical storage with backup and restore services** | **Allows basic sharing** |

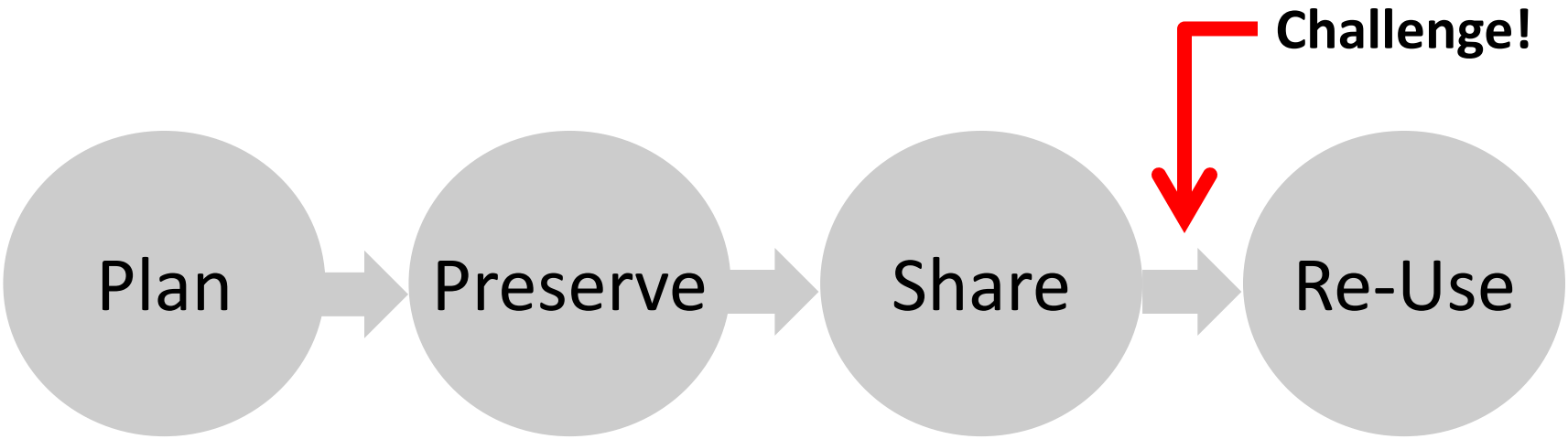
Choudhury, G. S., Palmer, C. L., Baker, K. S., & DiLauro, T. (2013, January). Levels of services and curation for high-functioning data. Presented at the International Digital Curation Conference, Amsterdam, Netherlands.



CC BY

# ACADIS: Scientific Data Preservation

*Challenge: Realizing curation level of service for research products*




| Collaboration | Data Provider | Expert Curators | Who? |
|---|---|---|---|
| Proposal Consulting | Self Publish Tools | Metadata QC | Data Standardization |
| DMP Template | REST Services | DOI Services | Rosetta Translator |
| Education | ISO 19115 Metadata | Curator Consulting | |
| | Chronopolis | Metadata Exchange | |

# Scientific Data Preservation

*Challenge: Realizing curation level of service for research products*


Options (require innovation):

- Motivate Data Providers
- Utilize Curation Staff
- Leverage Consumers

Data Providers != Data Consumers

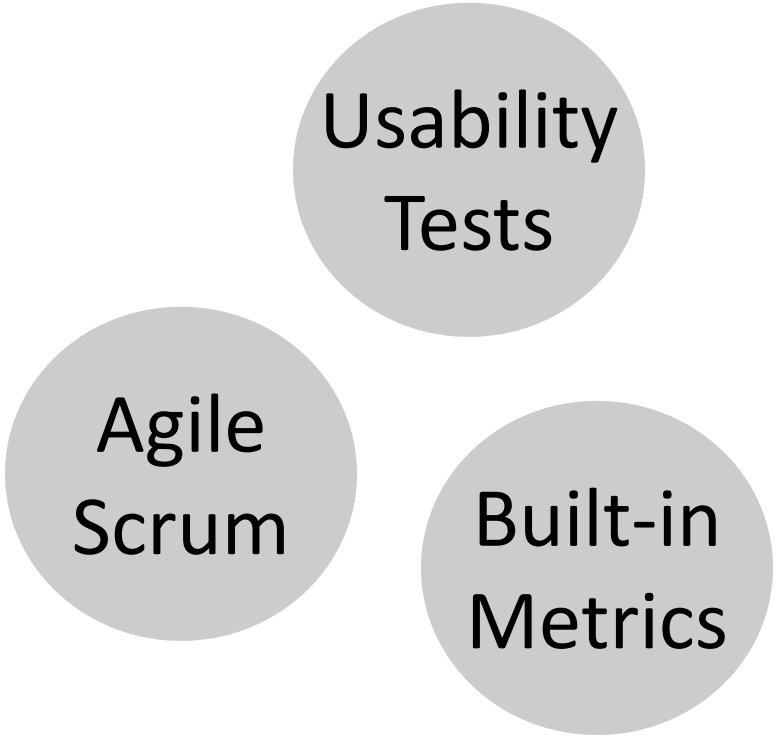
… and the gap is growing



CC BY

# ACADIS: User Centered Development

*Deliver high value first, test, measure and fail fast*





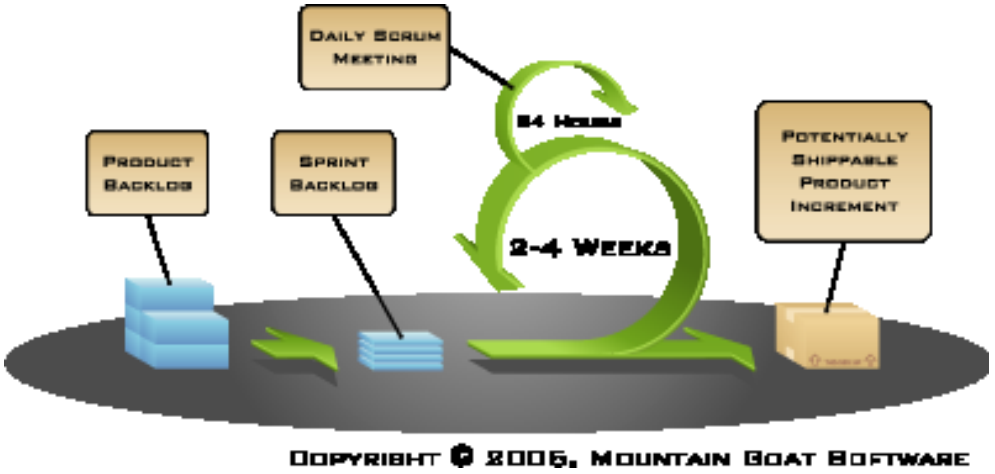

CC BY

# User Centered Development: Agile Scrum

*Deliver high value first in small batches…*

## Agile processes are transformative!

- Foundation for rapid delivery of **value**.
- Focus on **what is important** build, rather than “fun”.
- Release early, release often, **release well**.
- Value is from rigorous **attention to process**.

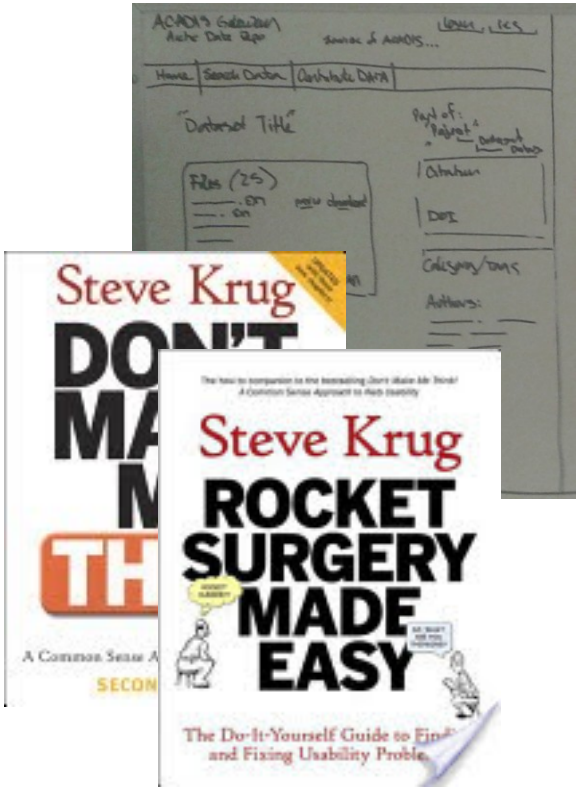


# User Centered Development: Usability Testing

*Low cost, high value learning*

## DIY Usability Tests are cheap with amazing return

- Do early, once a minimum viable feature is ready.
- Eye opening for developers (and product owner).
- Tests show overlooked **stumbling blocks**.
- Formal process key to gain maximum value.
- Paper prototypes work, too.
- May require consent forms.




A great usability focused site and blog:

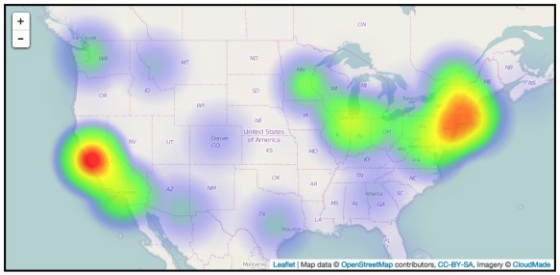
http://www.nngroup.com/

# User Centered Development: Measure Use

*Build metrics capture into your features to see what is actually used.*

## Metrics are for more than funding justification

Simple Mantra: Routinely look at metrics and logging.
Gather as much use data as possible.
Beware “vanity metrics” and focus on current use.
Build in metrics capture on new features.
Consider “A B” testing to compare features (we don’t do this yet.)
Also use the standard tools (http logs, Google Analytics, etc.)




# Wrap Up…

*Science portals require innovation, user centered processes and human experts*

- Testing and studying what we have is high value
- Embrace User Centered development processes
- Don’t be afraid to try and fail (as long as it is efficient)
- Focus curator experts in areas where technology falls short.

# Thank you!

ejn@ucar.edu

Funding: NSF Grant
Partners: NSIDC, NCAR: EOL, UCAR: Unidata



CC BY

Questions?

ejn@ucar.edu



CC BY

# Blank Slide

# Curators: Full Service Data Management

*Curation experts bridge certain gaps…*

- Full life cycle data management consulting.
- Data Management Plan template and guidance.
- Help desk system for communication and knowledge base.
- Direct input to product developers guides development.

## However, human experts are difficult to scale…